BLOCK-SPARSITY: COHERENCE AND EFFICIENT RECOVERY

Yonina C. Eldar

Technion, Haifa, Israel
yonina@ee.technion.ac.il

Helmut Bölcskei

ETH Zurich, Zurich, Switzerland
boelcskei@nari.ee.ethz.ch

ABSTRACT

We consider compressed sensing of block-sparse signals, i.e., sparse signals that have nonzero coefficients occuring in clusters. Based on an uncertainty relation for block-sparse signals, we define a block-coherence measure and we show that a block-version of the orthogonal matching pursuit algorithm recovers block k-sparse signals in no more than k steps if the block-coherence is sufficiently small. The same condition on block-sparsity is shown to guarantee successful recovery through a mixed ℓ_2/ℓ_1 optimization approach. The significance of the results lies in the fact that making explicit use of block-sparsity can yield better reconstruction properties than treating the signal as being sparse in the conventional sense thereby ignoring the additional structure in the problem.

Index Terms— block sparsity, coherence, uncertainty relations

1. INTRODUCTION

We consider compressed sensing [1, 2] of sparse signals that exhibit additional structure in the form of the nonzero coefficients occuring in clusters. It is therefore natural to ask whether explicitly taking this block sparse structure into account yields improvements over treating the signal as a conventional sparse signal. It was shown in [3, 4] that the answer is in the affirmative. Moreover, in [3] the restricted amplification property was shown to provide a sufficient condition for robust recovery of model-compressible (which includes block-sparse) signals. It is furthermore shown in [3] that simple modifications of the CoSaMP algorithm [5] and of iterative hard thresholding [6] yield reconstruction algorithms for the model-based case (including block-sparsity) that exhibit provable robustness properties. A mixed ℓ_2/ℓ_1-norm algorithm for recovering block-sparse signals was introduced in [4]. The block restricted isometry property defined in [4] provides equivalence conditions for guaranteeing recovery of block-sparse signals.

The focus of the present paper is on the notion of coherence for block-sparse signals, i.e., block-coherence, and can be seen as extending the program laid out in [7, 8] to the block-sparse case. We introduce a block version of the orthogonal matching pursuit algorithm (BOMP) and find a sufficient condition on block-coherence to guarantee recovery of block k-sparse signals through BOMP in no more than k steps. The same condition on block-coherence is shown to guarantee successful recovery through the mixed ℓ_2/ℓ_1 optimization approach, described in [4, 9]. These results are akin to a sufficient condition on conventional coherence reported in [7] that guarantees recovery through OMP or ℓ_1-optimization. Finally, we establish an uncertainty relation for block-sparse signals and show how the block-coherence measure defined previously occurs naturally in this uncertainty relation.

Notation. Throughout the paper, we denote vectors in $\mathbb{C}^N$ by boldface lowercase letters, e.g., $\mathbf{x}$, and matrices by boldface uppercase letters, e.g., $\mathbf{A}$. The identity matrix is written as $\mathbf{I}$ or $\mathbf{I}_d$ when the dimension is not clear from the context. Given a matrix $\mathbf{A}$, $\mathbf{A}^T$ and $\mathbf{A}^H$ are its transpose and conjugate transpose, respectively, $\mathbf{A}^\dagger$ is the pseudo inverse, $\mathcal{R}(\mathbf{A})$ denotes its range space, $\mathbf{A}_{i,j}$ is the element in the ith row and jth column, and $\mathbf{a}_\ell$ denotes its ℓth column. The ℓth element of a vector $\mathbf{x}$ is denoted by x_ℓ. The standard Euclidean norm is $\|\mathbf{x}\|_2 = \sqrt{\mathbf{x}^H\mathbf{x}}$, $\|\mathbf{x}\|_1 = \sum_\ell |x_\ell|$ is the ℓ_1-norm, $\|\mathbf{x}\|_\infty = \max_\ell |x_\ell|$ is the ℓ_∞-norm, and $\|\mathbf{x}\|_0$ designates the number of nonzero entries in $\mathbf{x}$. The Kronecker product of the matrices $\mathbf{A}$ and $\mathbf{B}$ is written as $\mathbf{A} \otimes \mathbf{B}$. The spectral radius of $\mathbf{A}$ is denoted as $\rho(\mathbf{A}) = \lambda_{\max}^{1/2}(\mathbf{A}^H\mathbf{A})$, where $\lambda_{\max}(\mathbf{B})$ is the largest eigenvalue of the positive-semidefinite matrix $\mathbf{B}$.

2. BLOCK-SPARSITY

Block-sparsity. We consider the problem of representing a vector $\mathbf{y} \in \mathbb{C}^L$ in a given dictionary $\mathbf{D}$ of size $L \times N$ with $L < N$, so that

$$\mathbf{y} = \mathbf{D}\mathbf{x} \tag{1}$$

for a coefficient vector $\mathbf{x} \in \mathbb{C}^N$. We require $\mathbf{x}$ to be block-sparse, where, throughout the paper, blocks are always assumed to be of length d. To define block-sparsity, we view $\mathbf{x}$ as a concatenation of blocks (of length d) with $\mathbf{x}[\ell]$ denoting the ℓth sub-block, i.e.,

$$\mathbf{x}^T = [\underbrace{x_1 \ \ldots \ x_d}_{\mathbf{x}[1]} \ \underbrace{x_{d+1} \ \ldots x_{2d}}_{\mathbf{x}[2]} \ \cdots \ \underbrace{x_{N-d+1} \ \ldots \ x_N}_{\mathbf{x}[M]}]^T \tag{2}$$

with $N = Md$. We furthermore assume that $L = Rd$ with R integer. A vector $\mathbf{x} \in \mathbb{C}^N$ is called block k-sparse if $\mathbf{x}[\ell]$ has nonzero Euclidean norm for at most k indices ℓ. When $d = 1$, block-sparsity reduces to the conventional definition of sparsity as in [1, 2]. Denoting

$$\|\mathbf{x}\|_{2,0} = \sum_{\ell=1}^{M} I(\|\mathbf{x}[\ell]\|_2 > 0) \tag{3}$$

where $I(\|\mathbf{x}[\ell]\|_2 > 0) = 1$ if $\|\mathbf{x}[\ell]\|_2 > 0$ and 0 otherwise, a block k-sparse vector $\mathbf{x}$ is defined as a vector that satisfies $\|\mathbf{x}\|_{2,0} \le k$. In the remainder of the paper conventional sparsity will be referred to simply as sparsity, in contrast to block-sparsity.

Problem statement. Our goal is to provide conditions on the dictionary $\mathbf{D}$ ensuring that the block-sparse vector $\mathbf{x}$ can be reconstructed from measurements of the form (1) through computationally efficient algorithms. Our approach is largely based on [7, 10] (and the mathematical techniques used therein) where equivalent results are provided for the sparse case. The results in [7, 10] are stated in terms of the dictionary coherence. Therefore, as a first step in our development, we extend this conventional coherence measure to block-sparsity by defining block-coherence. Before introducing the corresponding definition, we cite the following proposition taken from [4].

Proposition 1. *The representation (1) is unique if and only if* $\mathbf{Dg} \neq \mathbf{0}$ *for every* $\mathbf{g} \neq \mathbf{0}$ *that is block* $2k$*-sparse.*

Similarly to (2), we can represent $\mathbf{D}$ as a concatenation of column-blocks $\mathbf{D}[\ell]$ of size $L \times d$:

$$\mathbf{D} = [\underbrace{\mathbf{d}_1 \ \dots \ \mathbf{d}_d}_{\mathbf{D}[1]} \ \underbrace{\mathbf{d}_{d+1} \ \dots \mathbf{d}_{2d}}_{\mathbf{D}[2]} \ \dots \ \underbrace{\mathbf{d}_{N-d+1} \ \dots \ \mathbf{d}_N}_{\mathbf{D}[M]}]. \quad (4)$$

Since from Proposition 1 the columns of $\mathbf{D}[\ell]$, $\forall \ell$, are linearly independent, we may write $\mathbf{D}[\ell] = \mathbf{A}[\ell]\mathbf{W}_\ell$ where $\mathbf{A}[\ell]$ consists of orthonormal columns that span $\mathcal{R}(\mathbf{D}[\ell])$ and $\mathbf{W}_\ell$ is invertible. Denoting by $\mathbf{A}$ the $L \times N$ matrix with blocks $\mathbf{A}[\ell]$, and by $\mathbf{W}$ the $N \times N$ block-diagonal matrix with blocks $\mathbf{W}_\ell$, we conclude that $\mathbf{D} = \mathbf{AW}$. Since $\mathbf{W}$ is block-diagonal and invertible, $\mathbf{c} = \mathbf{Wx}$ is block-sparse with the same block-sparsity level as $\mathbf{x}$. Therefore, in the sequel, we assume, without loss of generality, that $\mathbf{D}$ consists of orthonormal blocks, i.e., $\mathbf{D}^H[\ell]\mathbf{D}[\ell] = \mathbf{I}_d$. Throughout the paper, we furthermore assume that the dictionaries we consider satisfy the condition of Proposition 1.

Block-coherence. We define the block-coherence of $\mathbf{D}$ as

$$\mu_{\mathrm{B}} = \max_{\ell, r \neq \ell} \frac{1}{d} \rho(\mathbf{M}[\ell, r]) \quad \text{with} \quad \mathbf{M}[\ell, r] = \mathbf{D}^H[\ell]\mathbf{D}[r]. \quad (5)$$

Note that $\mathbf{M}[\ell, r]$ is the ℓrth $d \times d$ block of the $N \times N$ matrix $\mathbf{M} = \mathbf{D}^H\mathbf{D}$. When $d = 1$, μ_{B} reduces to the conventional definition of coherence [11, 10, 7]

$$\mu = \max_{\ell, r \neq \ell} |\mathbf{d}_\ell^H \mathbf{d}_r|. \quad (6)$$

It is easy to see that the definition in (5) is invariant to the choice of orthonormal basis $\mathbf{D}[\ell]$ for $\mathcal{R}(\mathbf{D}[\ell])$. This is because $\rho(\mathbf{M}[\ell, r]) = \rho(\mathbf{U}_\ell^H \mathbf{M}[\ell, r]\mathbf{U}_r)$. In the remainder of the paper conventional coherence will be referred to simply as coherence, in contrast to block-coherence.

Proposition 2. *The block-coherence* μ_{B} *satisfies* $0 \leq \mu_{\mathrm{B}} \leq 1$.

Proof. Clearly $\mu_{\mathrm{B}} \geq 0$. To prove that $\mu_{\mathrm{B}} \leq 1$, note that $\rho(\mathbf{A}) \leq \|\mathbf{A}\|$, where $\|\mathbf{A}\|$ is any matrix norm. In particular, if $\mathbf{A}$ is a $d \times d$ matrix, then

$$\rho(\mathbf{A}) \leq \max_j \sum_i |\mathbf{A}_{i,j}| \leq d \max_{i,j} |\mathbf{A}_{i,j}|. \quad (7)$$

In our case, $\mathbf{A} = \mathbf{M}[\ell, r]$. Since the columns of $\mathbf{D}$ are normalized, all the elements of $\mathbf{M}[\ell, r]$ have absolute value smaller than or equal to 1, so that from (7), $\rho(\mathbf{M}[\ell, r]) \leq d$, and hence $\mu_{\mathrm{B}} \leq 1$. $\square$

It is interesting to compare μ_{B} with the coherence μ defined in (6) for the same dictionary $\mathbf{D}$.

Proposition 3. *For any dictionary* $\mathbf{D}$*, we have* $\mu_{\mathrm{B}} \leq \mu$.

The proof follows immediately from (7).

3. UNCERTAINTY RELATION FOR BLOCK-SPARSITY

We next show how the block-coherence μ_{B} defined above naturally appears in an uncertainty relation for block-sparse signals. This uncertainty relation generalizes the corresponding result for the sparse case reported in [10].

The uncertainty principle for the sparse case is concerned with pairs of representations of a vector $\mathbf{x} \in \mathbb{C}^N$ in two different orthonormal bases for $\mathbb{C}^N$: $\{\boldsymbol{\phi}_\ell, 1 \leq \ell \leq N\}$ and $\{\boldsymbol{\psi}_\ell, 1 \leq \ell \leq N\}$ [11, 10]. Any vector $\mathbf{x} \in \mathbb{C}^N$ can be expanded uniquely in terms of each one of these bases according to:

$$\mathbf{x} = \sum_{\ell=1}^{N} a_\ell \boldsymbol{\phi}_\ell = \sum_{\ell=1}^{N} b_\ell \boldsymbol{\psi}_\ell. \quad (8)$$

The uncertainty relation sets limits on the sparsity of the decompositions (8) for any $\mathbf{x} \in \mathbb{C}^N$. Specifically, denoting $A = \|\mathbf{a}\|_0$ and $B = \|\mathbf{b}\|_0$, it is shown in [10] that

$$\frac{1}{2}(A + B) \geq \sqrt{AB} \geq \frac{1}{\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})} \quad (9)$$

where $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ is the coherence between $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, defined by

$$\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi}) = \max_{\ell, r} |\boldsymbol{\phi}_\ell^H \boldsymbol{\psi}_r|. \quad (10)$$

In [11] it is shown that $1/\sqrt{N} \leq \mu(\boldsymbol{\Phi}, \boldsymbol{\Psi}) \leq 1$. We now develop an uncertainty principle for block-sparse decompositions, analogous to (9). Specifically, we find a result that is equivalent to (9) with A and B replaced by block-sparsity levels as defined in (3) and $\mu(\boldsymbol{\Phi}, \boldsymbol{\Psi})$ replaced by the block-coherence between the orthonormal bases considered, as defined in (13).

Theorem 1. *[12] Let* $\boldsymbol{\Phi}, \boldsymbol{\Psi}$ *be two unitary matrices with* $L \times d$ *blocks* $\{\boldsymbol{\Phi}[\ell], \boldsymbol{\Psi}[\ell], 1 \leq \ell \leq M\}$ *and let* $\mathbf{x} \in \mathbb{C}^N$ *satisfy*

$$\mathbf{x} = \sum_{\ell=1}^{M} \boldsymbol{\Phi}[\ell]\mathbf{a}[\ell] = \sum_{\ell=1}^{M} \boldsymbol{\Psi}[\ell]\mathbf{b}[\ell]. \quad (11)$$

Let $A = \|\mathbf{a}\|_{2,0}$ *and* $B = \|\mathbf{b}\|_{2,0}$*. Then,*

$$\frac{1}{2}(A + B) \geq \sqrt{AB} \geq \frac{1}{d\mu_{\mathrm{B}}(\boldsymbol{\Phi}, \boldsymbol{\Psi})} \quad (12)$$

where

$$\mu_{\mathrm{B}}(\boldsymbol{\Phi}, \boldsymbol{\Psi}) = \max_{\ell, r} \frac{1}{d} \rho(\boldsymbol{\Phi}^H[\ell]\boldsymbol{\Psi}[r]). \quad (13)$$

It can easily be shown that for $\mathbf{D}$ consisting of the orthonormal bases $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, i.e., $\mathbf{D} = [\boldsymbol{\Phi} \ \boldsymbol{\Psi}]$, we have $\mu_{\mathrm{B}}(\boldsymbol{\Phi}, \boldsymbol{\Psi}) = \mu_{\mathrm{B}}$, where μ_{B} is as defined in (5) and associated with $\mathbf{D} = [\boldsymbol{\Phi} \ \boldsymbol{\Psi}]$.

The bound provided by Theorem 1 can be tighter than that obtained by applying the conventional uncertainty relation (9) to the block-sparse case. This can be seen by using $\|\mathbf{a}\|_0 \leq d\|\mathbf{a}\|_{2,0}$, $\|\mathbf{b}\|_0 \leq d\|\mathbf{b}\|_{2,0}$, and (9) to obtain

$$\sqrt{\|\mathbf{a}\|_{2,0}\|\mathbf{b}\|_{2,0}} \geq \frac{1}{d\mu}. \quad (14)$$

Since $\mu_{\mathrm{B}} \leq \mu$, this bound can be looser than (12).

3.1. Block-incoherent dictionaries

As already noted, in the sparse case (i.e., $d = 1$) for any two orthonormal bases $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, we have $\mu \geq 1/\sqrt{N}$. We next show that the block-coherence satisfies a similar inequality, namely $\mu_{\mathrm{B}} \geq 1/\sqrt{dN}$. Evidently, the lower bound on μ is $\sqrt{d}$ times larger than that on μ_{B}. To prove the lower bound on μ_{B}, let $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ denote two orthonormal bases for $\mathbb{C}^N$ and let $\mathbf{A} = \boldsymbol{\Phi}^H\boldsymbol{\Psi}$ where $\mathbf{A}[\ell, r]$ stands for the (ℓ, r)th $d \times d$ block of $\mathbf{A}$. With $M = N/d$, we have

$$\begin{aligned} M^2\mu_{\mathrm{B}}^2 &\geq \sum_{\ell=1}^{M}\sum_{r=1}^{M} \frac{1}{d^2} \lambda_{\max}(\mathbf{A}^H[\ell, r]\mathbf{A}[\ell, r]) \\ &\geq \frac{1}{d^2} \lambda_{\max} \left(\sum_{\ell=1}^{M}\sum_{r=1}^{M} \mathbf{A}^H[\ell, r]\mathbf{A}[\ell, r] \right). \quad (15) \end{aligned}$$

Now, it holds that

$$\sum_{\ell=1}^{M}\sum_{r=1}^{M}\mathbf{A}^H[\ell,r]\mathbf{A}[\ell,r]=\sum_{r=1}^{M}\mathbf{\Psi}^H[r]\left(\sum_{\ell=1}^{M}\mathbf{\Phi}[\ell]\mathbf{\Phi}^H[\ell]\right)\mathbf{\Psi}[r]. \tag{16}$$

Since $\mathbf{\Phi}$ consists of orthonormal columns, $\sum_\ell \mathbf{\Phi}[\ell]\mathbf{\Phi}^H[\ell] = \mathbf{\Phi}\mathbf{\Phi}^H = \mathbf{I}_L$. Furthermore, since $\mathbf{\Psi}[r]$ consists of orthonormal columns, $\forall r$, we have $\mathbf{\Psi}^H[r]\mathbf{\Psi}[r]=\mathbf{I}_d$, $\forall r$. Therefore, (15) becomes

$$\mu_{\rm B}^2 \geq \frac{1}{Md^2}=\frac{1}{dN} \tag{17}$$

which concludes the proof.

We now construct a pair of bases that achieves the lower bound on $\mu_{\rm B}$ and therefore has the smallest possible block-coherence. Let $\mathbf{F}$ be the DFT matrix of size $M = N/d$ with $\mathbf{F}_{\ell,r} = (1/\sqrt{M})\exp(j2\pi\ell r/M)$. Define $\mathbf{\Phi}=\mathbf{I}_N$ and

$$\mathbf{\Psi}=\mathbf{F}\otimes\mathbf{U}_d \tag{18}$$

where $\mathbf{U}_d$ is an arbitrary $d\times d$ unitary matrix. For this choice, $\mathbf{\Phi}^H[\ell]\mathbf{\Psi}[r]=\mathbf{F}_{\ell,r}\mathbf{U}_d$. Since $\rho(\mathbf{U}_d)=1$ and $|\mathbf{F}_{\ell,r}|=1/\sqrt{M}$, we get

$$\mu_{\rm B}=\frac{1}{d\sqrt{M}}=\frac{1}{\sqrt{dN}}. \tag{19}$$

When $d=1$, this basis pair reduces to the spike-Fourier pair which is well known to be maximally incoherent [11].

4. EFFICIENT RECOVERY ALGORITHMS

We now give operational meaning to block-coherence by showing that if it is small enough, a block-sparse signal $\mathbf{x}$ can be recovered from $\mathbf{y}=\mathbf{D}\mathbf{x}$ using computationally efficient algorithms. We consider two different algorithms, namely the mixed ℓ_2/ℓ_1 optimization program proposed in [4]:

$$\min_{\mathbf{x}}\sum_{\ell=1}^{M}\|\mathbf{x}[\ell]\|_2 \quad \mathrm{s.t.}\ \mathbf{y}=\mathbf{D}\mathbf{x} \tag{20}$$

and an extension of the orthogonal matching pursuit (OMP) algorithm [13] to the block-sparse case described below and termed BOMP. We then show that both methods recover the correct block-sparse $\mathbf{x}$ as long as $\mu_{\rm B}$ associated with $\mathbf{D}$ is small enough.

4.1. Block OMP

The BOMP algorithm is similar in spirit to the conventional OMP algorithm, and can serve as a computationally attractive alternative to (20).

The algorithm begins by initializing the residual as $\mathbf{r}_0=\mathbf{y}$. At the ℓth stage ($\ell\geq 1$) we choose the subspace that is best matched to $\mathbf{r}_{\ell-1}$ according to:

$$i_\ell=\arg\max\|\mathbf{D}^H[i]\mathbf{r}_{\ell-1}\|_2. \tag{21}$$

Once the index i_ℓ is chosen, we find the optimal coefficients by computing $\mathbf{x}_\ell[i]$ as the solution to

$$\min\left\|\mathbf{y}-\sum_{i\in\mathcal{I}}\mathbf{D}[i]\mathbf{x}_\ell[i]\right\|_2^2. \tag{22}$$

Here $\mathcal{I}$ is the set of chosen indices $i_j, 1\leq j\leq \ell$. The residual is then updated as

$$\mathbf{r}_\ell=\mathbf{y}-\sum_{i\in\mathcal{I}}\mathbf{D}[i]\mathbf{x}_\ell[i]. \tag{23}$$

4.2. Recovery conditions

Our main result, summarized in Theorem 3 below, is that any block k-sparse vector $\mathbf{x}$ can be recovered from measurements $\mathbf{y}=\mathbf{D}\mathbf{x}$ using either the BOMP algorithm or (20) if the block-coherence satisfies $kd<(\mu_{\rm B}^{-1}+d)/2$. If $\mathbf{x}$ was treated as a (conventional) kd-sparse vector without exploiting knowledge of the block-sparse structure, a sufficient condition for perfect recovery using OMP or (20) for $d=1$ (a.k.a. basis pursuit) is $kd<(\mu^{-1}+1)/2$. Since $\mu\geq\mu_{\rm B}$, exploiting the block structure by using BOMP or (20) recovery is guaranteed for a potentially higher sparsity level.

To state our results, suppose that $\mathbf{x}_0$ is a length-N block k-sparse vector, and let $\mathbf{y}=\mathbf{D}\mathbf{x}_0$ where $\mathbf{D}$ consists of blocks $\mathbf{D}[\ell]$ with orthonormal columns. Let $\mathbf{D}_0$ denote the $L\times(kd)$ matrix whose blocks correspond to the non-zero blocks of $\mathbf{x}_0$, and let $\overline{\mathbf{D}}_0$ be the matrix of size $L\times(N-kd)$ which contains the columns of $\mathbf{D}$ not in $\mathbf{D}_0$. We then have the following theorem proved in Section 5.

Theorem 2. *Let $\mathbf{x}_0\in\mathbb{C}^N$ be a block k-sparse vector with blocks of length d, and let $\mathbf{y}=\mathbf{D}\mathbf{x}_0$ for a given $L\times N$ matrix $\mathbf{D}$. A sufficient condition for the output of the BOMP and of (20) to equal $\mathbf{x}_0$ is that*

$$\rho_c(\mathbf{D}_0^\dagger\overline{\mathbf{D}}_0)<1 \tag{24}$$

where

$$\rho_c(\mathbf{A})=\max_\ell\sum_r\rho(\mathbf{A}[r,\ell]) \tag{25}$$

and $\mathbf{A}[r,\ell]$ is the (r,ℓ)th $d\times d$ block of $\mathbf{A}$.

Note that

$$\rho_c(\mathbf{D}_0^\dagger\overline{\mathbf{D}}_0)=\max_\ell\rho_c(\mathbf{D}_0^\dagger\overline{\mathbf{D}}_0[\ell]). \tag{26}$$

Therefore, (24) implies that for all ℓ,

$$\rho_c(\mathbf{D}_0^\dagger\overline{\mathbf{D}}_0[\ell])<1. \tag{27}$$

The sufficient condition (24) depends on $\mathbf{D}_0$ and hence on the location of the nonzero blocks in $\mathbf{x}_0$, which, of course, is not known in advance. Nonetheless, as the following theorem shows, (24) holds whenever the dictionary $\mathbf{D}$ has low block-coherence.

Theorem 3. *[12] Let $\mu_{\rm B}$ be the block-coherence defined by (5). Then (24) is satisfied if*

$$kd<\frac{1}{2}(\mu_{\rm B}^{-1}+d). \tag{28}$$

For $d=1$, we recover the results of [7, 8].

5. PROOF OF THEOREM 2

We start with some definitions. For $\mathbf{x}\in\mathbb{C}^N$, we define the general mixed ℓ_2/ℓ_p norm:

$$\|\mathbf{x}\|_{2,p}=\|\mathbf{v}\|_p,\quad \text{where } v_\ell=\|\mathbf{x}[\ell]\|_2, \tag{29}$$

and the $\mathbf{x}[\ell]$ are consecutive length-d blocks. For an $L\times N$ matrix $\mathbf{A}$ with $L=Rd$ and $N=Md$, where R and M are integers, we define the mixed matrix norm (with block size d) as

$$\|\mathbf{A}\|_{2,p}=\max_{\mathbf{x}}\frac{\|\mathbf{A}\mathbf{x}\|_{2,p}}{\|\mathbf{x}\|_{2,p}}. \tag{30}$$

The following lemma provides bounds on the mixed matrix norms for $p=1,\infty$, which we will use in the sequel.

Lemma 1. *[12] Let $\mathbf{A}$ be an $L \times N$ matrix with $L = Rd$ and $N = Md$. Denote by $\mathbf{A}[\ell, r]$ the (ℓ, r)th $d \times d$ block of $\mathbf{A}$. Then,*

$$\|\mathbf{A}\|_{2,\infty} \leq \max_r \sum_\ell \rho(\mathbf{A}[r,\ell]) \stackrel{\triangle}{=} \rho_r(\mathbf{A}) \tag{31}$$

$$\|\mathbf{A}\|_{2,1} \leq \max_\ell \sum_r \rho(\mathbf{A}[r,\ell]) \stackrel{\triangle}{=} \rho_c(\mathbf{A}). \tag{32}$$

In particular, $\rho_r(\mathbf{A}) = \rho_c(\mathbf{A}^H)$.

5.1. Block OMP

We begin by proving that (24) is sufficient to ensure recovery using the BOMP algorithm.

To prove the result, we first show that if $\mathbf{r}_{\ell-1}$ is in $\mathcal{R}(\mathbf{D}_0)$, then the next chosen index i_ℓ will be correct, namely it will correspond to a block in $\mathbf{D}_0$. Assuming that this is true, it follows immediately that i_1 is correct since clearly $\mathbf{r}_0 = \mathbf{y}$ lies in $\mathcal{R}(\mathbf{D}_0)$. Noting that $\mathbf{r}_\ell$ lies in the space spanned by $\mathbf{y}$ and $\mathbf{D}_0[i], i \in \mathcal{I}_\ell$, where $\mathcal{I}_\ell$ denotes the indices chosen up to stage ℓ, it follows that if $\mathcal{I}_\ell$ corresponds to correct indices, i.e., $\mathbf{D}[i]$ is a block of $\mathbf{D}_0$ for all $i \in \mathcal{I}_\ell$, then $\mathbf{r}_\ell$ also lies in $\mathcal{R}(\mathbf{D}_0)$ and the next index will be correct as well. Thus, at every step a correct subset is selected. It is also clear that no index will be chosen twice since the new residual is orthogonal to all the previously chosen subspaces; consequently the correct $\mathbf{x}_0$ will be recovered in k steps.

It therefore remains to show that if $\mathbf{r}_{\ell-1} \in \mathcal{R}(\mathbf{D}_0)$, then under (24) the next chosen index corresponds to a block in $\mathbf{D}_0$. This is equivalent to requiring that

$$z(\mathbf{r}_{\ell-1}) = \frac{\|\overline{\mathbf{D}}_0^H \mathbf{r}_{\ell-1}\|_{2,\infty}}{\|\mathbf{D}_0^H \mathbf{r}_{\ell-1}\|_{2,\infty}} < 1. \tag{33}$$

From the properties of the pseudo-inverse, $\mathcal{R}(\mathbf{D}_0) = \mathcal{R}(\mathbf{D}_0\mathbf{D}_0^\dagger)$, and consequently $\mathbf{D}_0\mathbf{D}_0^\dagger \mathbf{r}_{\ell-1} = \mathbf{r}_{\ell-1}$. Since $\mathbf{D}_0\mathbf{D}_0^\dagger$ is Hermitian,

$$(\mathbf{D}_0^\dagger)^H \mathbf{D}_0^H \mathbf{r}_{\ell-1} = \mathbf{r}_{\ell-1}. \tag{34}$$

Substituting (34) into (33) yields $z(\mathbf{r}_{\ell-1}) =$

$$\frac{\|\overline{\mathbf{D}}_0^H (\mathbf{D}_0^\dagger)^H \mathbf{D}_0^H \mathbf{r}_{\ell-1}\|_{2,\infty}}{\|\mathbf{D}_0^H \mathbf{r}_{\ell-1}\|_{2,\infty}} \leq \rho_r(\overline{\mathbf{D}}_0^H (\mathbf{D}_0^\dagger)^H) = \rho_c(\mathbf{D}_0^\dagger \overline{\mathbf{D}}_0), \tag{35}$$

where we used Lemma 1. This completes the proof.

5.2. ℓ_2/ℓ_1 Optimization

We now show that (24) is also sufficient to ensure recovery using (20). To this end we rely on the following lemma:

Lemma 2. *[12] Suppose that $\mathbf{v}$ is a length $N = Md$ vector with $\|\mathbf{v}[\ell]\|_2 > 0, \forall l$, and that $\mathbf{A}$ is a matrix of size $L \times N$, where $L = Rd$ and the blocks $\mathbf{A}[\ell, r]$ are of size $d \times d$. Then, $\|\mathbf{Av}\|_{2,1} \leq \rho_c(\mathbf{A})\|\mathbf{v}\|_{2,1}$. If in addition the values of $\rho_c(\mathbf{AJ}_\ell)$ are not all equal, then the inequality is strict. Here, $\mathbf{J}_\ell$ is an $N \times d$ matrix that is all zero except for the ℓth $d \times d$ block which equals $\mathbf{I}_d$.*

To prove that (20) recovers the correct vector $\mathbf{x}_0$, let $\mathbf{x}'$ be another set of coefficients for which $\mathbf{y} = \mathbf{Dx}'$. Denote by $\mathbf{c}_0$ and $\mathbf{c}'$ the length kd vectors consisting of the non-zero elements of $\mathbf{x}_0$ and $\mathbf{x}'$, respectively. Let $\mathbf{D}_0$ and $\mathbf{D}'$ denote the corresponding columns of $\mathbf{D}$ so that $\mathbf{y} = \mathbf{D}_0\mathbf{c}_0 = \mathbf{D}'\mathbf{c}'$. From the assumption in Proposition 1, it follows that there cannot be two different representations using the same blocks $\mathbf{D}_0$. Therefore, $\mathbf{D}'$ must contain at least one block, $\mathbf{Z}$, that is not included in $\mathbf{D}_0$. From (27), $\rho_c(\mathbf{D}_0^\dagger \mathbf{Z}) < 1$. For any other block $\mathbf{U}$ in $\mathbf{D}$, we must have that

$$\rho_c(\mathbf{D}_0^\dagger \mathbf{U}) \leq 1. \tag{36}$$

Indeed, if $\mathbf{U} \in \mathbf{D}_0$, then $\mathbf{U} = \mathbf{D}_0[\ell] = \mathbf{D}_0\mathbf{J}_\ell$ where $\mathbf{J}_\ell$ is a matrix with d columns which is all zero, except for the ℓth block which is equal to $\mathbf{I}_d$. In this case, $\mathbf{D}_0^\dagger\mathbf{D}_0[\ell] = \mathbf{J}_\ell$ and hence $\rho_c(\mathbf{D}_0^\dagger\mathbf{D}_0[l]) = \rho_c(\mathbf{D}_0^\dagger\mathbf{U}) = 1$. If, on the other hand, $\mathbf{U} = \overline{\mathbf{D}}[\ell]$ for some ℓ, then it follows from (27) that $\rho_c(\mathbf{D}_0^\dagger\mathbf{U}) < 1$.

Now, suppose first that the blocks in $\mathbf{D}_0^\dagger\mathbf{D}'$ do not all have the same spectral radius ρ. Then,

$$\begin{aligned} \|\mathbf{c}_0\|_{2,1} &= \|\mathbf{D}_0^\dagger\mathbf{D}_0\mathbf{c}_0\|_{2,1} = \|\mathbf{D}_0^\dagger\mathbf{y}\|_{2,1} = \|\mathbf{D}_0^\dagger\mathbf{D}'\mathbf{c}'\|_{2,1} \\ &< \rho_c(\mathbf{D}_0^\dagger\mathbf{D}')\|\mathbf{c}'\|_{2,1} \leq \|\mathbf{c}'\|_{2,1} \end{aligned} \tag{37}$$

where the first equality stems from the fact that the columns of $\mathbf{D}_0$ are linearly independent (a consequence of the assumption in Proposition 1), the first inequality follows from Lemma 2 since $\|\mathbf{c}'[\ell]\|_2 > 0$, $\forall l$, and the last inequality follows from (36). If all the blocks of $\mathbf{D}_0^\dagger\mathbf{D}'$ have identical spectral radius ρ, then $\rho < 1$ as for $\mathbf{Z} \in \mathbf{D}'$, $\rho_c(\mathbf{D}_0^\dagger\mathbf{Z}) < 1$. Repeating the calculations in (37), we find that the first inequality is no longer strict. However, the second inequality in (37) is strict instead so that the conclusion still holds.

Since $\|\mathbf{x}_0\|_{2,1} = \|\mathbf{c}_0\|_{2,1}$ and $\|\mathbf{x}'\|_{2,1} = \|\mathbf{c}'\|_{2,1}$, we conclude that under (27), any set of coefficients used to represent the original signal that is not equal to $\mathbf{x}_0$ will result in a larger ℓ_2/ℓ_1 norm.

6. REFERENCES

[1] E. J. Candes, J. K. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[2] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[3] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, “Model-based compressive sensing,” *IEEE Trans. Inf. Theory*, 2008, submitted.

[4] Y. C. Eldar and M. Mishali, “Robust recovery of signals from a union of subspaces,” *IEEE Trans. Inf. Theory*, 2008, submitted.

[5] D. Needell and J. A. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” *Applied and Computational Harmonic Analysis*, June 2008, accepted for publication.

[6] T. Blumensath and M. E. Davies, “Iterative hard thresholding for compressed sensing,” May 2008, submitted.

[7] J. Tropp, “Greed is good: Algorithmic results for sparse approximation,” *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[8] D. L. Donoho and M. Elad, “Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization,” *Proc. Natl. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, March 2003.

[9] M. Stojnic, F. Parvaresh, and B. Hassibi, “On the reconstruction of block-sparse signals with an optimal number of measurements,” April 2008, submitted.

[10] M. Elad and A. M. Bruckstein, “A generalized uncertainty principle and sparse representation in pairs of bases,” *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2558–2567, Sept. 2002.

[11] D. L. Donoho and X. Huo, “Uncertainty principles and ideal atomic decompositions,” *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[12] Y. C. Eldar and H. Bölcskei, “Compressed sensing for block sparse signals: Uncertainty relations, coherence, and efficient recovery,” 2008, in preparation.

[13] G. Davis, S. Mallat, and Z. Zhang, “Adaptive time-frequency decompositions,” *Optical Engineering*, vol. 33, no. 7, pp. 2183–2191, July 1994.